Maximum Likelihood Rover Localization by Matching Range Maps

Clark F. Olson
Jet Propulsion Laboratory, California Institute of Technology
Mail Stop 107-1 W?, 4800 Oak Grove Drive, Pasadena, CA 91109
http: //robotics. jpl.nasa.gov/people/olson/homepage.html

Abstract

This paper describes maximum likelihood estimation techniques for performing rover localization in natural terrain by matching range maps. An occupancy map of the local terrain is generated using stereo vision. The position of the rover with respect to a previously generated occupancy map is then computed by comparing the maps using a probabilistic formulation of Hausdorff matching techniques. Our motivation for this work is the desire for greater autonomy in Mars rovers. These techniques have been applied to data obtained from the Sojourner Mars rover and ran onboard the Rocky 7 Mars rover prototype.

1 Introduction

Visual sensors can be used to reduce the positional uncertainty in mobile robots that is accumulated due to dead-reckoning error [13]. This paper describes a method for performing self-localization in natural terrain by matching a range map generated from the robot cameras (the local *map*) to a range map encompassing the same terrain that has been previously generated (the *global map).*

10 perform localization, we first compute an occupancy map of the terrain from stereo imagery. This local map is compared to the global map using a similarity measure based on the Hausdorff distance [5]. We have generalized the measure to arbitrary probability distributions using the principal of maximum likelihood estimation and modified it to apply to terrain occupancy maps. The best relative position between the maps is located using a hierarchical search that guarantees that the best position in a discretized pose space is found.

Our motivation for pursuing this work is the Long Range Science Rover project at J PL, which has developed the Rocky 7 Mars rover prototype [4]. There is a current need for increased self-localization ability in Mars rovers in order to perform with greater autonomy from both operators on Earth and from the lander bringing the rover to Mars. For example, the Sojourner Mars rover is limited to moving short distances during each downlink cycle due to positional uncertainty, The method by which dead-reckoning errors are corrected for Sojourner is through a human operator overlaying a model of the rover on stereo range data computed from downlinked imagery of the rover taken by the lander [12].

'J'here are a number of scenarios in which these techniques can be used in the context of a Mars mission. While operating in a small area containing several science targets (such as the area around the lander that Sojourner has operated in), we may perform localization using the panoramic imagery generated at the center of the area as our global map. While this is not crucial when the lander can see the rover. The next generation Mars rover will have a mast with stereo cameras that will allow it to generate panoramic imagery when it is not near the lander. See Figure 1. We can then perform localization by matching the panoramic range maps generated using the mast imagery to maps generated from either the low-to-the-ground navigation cameras, if possible, or by raising the mast to image interesting terrain, if necessary.

Previous work on performing localization using visual landmarks for rovers on extra-terrestrial missions [1] has concentrated on rough localization in a large area by detecting mountain peaks and maximizing the posterior probability of the position given the directions to the mountains. The average error in the localization of this system is 91 meters in the two experiments reported. In contrast, we are concerned with fine localization in a relatively small area, and achieve errors much smaller than a meter.

The techniques described here are effective whenever a dense range map can be generated in the robot local coordinate frame and we have a range map of tile same terrain in the frame of reference in which we wish to localize the robot. We can thus use either

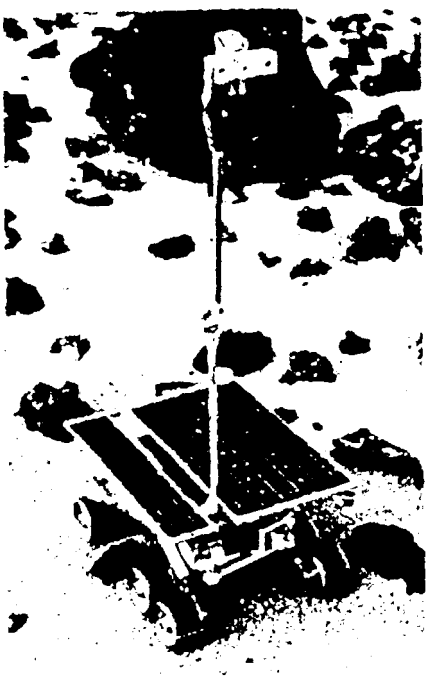


Figure 1: Rocky 7 Mars rover prototype in the JPL Mars yard with mast deployed.

rover mast or navigation imagery to generate the local map. The globs] map might also consist of the rover mast or navigation imagery, but it could also consist of imagery from the lander (including descent imagery), and it is possible that orbital imagery could be used, although it is not likely that we will have orbital imagery of sufficient resolution to use for accurate rover localization in the near future [8].

We have tested these techniques by matching terrain on Liars that was imaged with Sojourner's stereo cameras to a terrain map generated from the stereo cameras on the Pathfinder lander. The results indicate that self-localization can performed with these techniques nearly as well as a human operator without requiring a downlink cycle. These techniques have also been implemented on-board Rocky 7 with good results.

2 Terrain maps

While we could potentially use any method for generating three-dimensional range data of the terrain, we concentrate on the use of stereo vision, since this is the method by which Rocky 7 computes range maps for obstacle detection. The techniques that we use to compute the stereo range data have been described elsewhere [6, 1'], so we summarize only tile important points here.

An off-line step, where the stereo camera rig is calibrated, must first be performed. We use a camera model that allows arbitrary affine transformation of the image plane [14] and that has been extended to include radial lens distortion [2]. The remainder of the method is performed on-line.

At run-time, each image first warped to remove the lens distortion and the images are rectified so that the corresponding scan-lines yield corresponding epipolar lines in the image. The disparity between the left and right images is measured for each pixel by minimizing the sum-of-squared-difference (SSD) measure of windows around the pixel in the Laplacian of the image over a finite disparity range. Subpixel disparity estimates are computed using parabolic interpolation on tile SSD values neighboring the minimum. Outliers are removed through consistency checking and smoothing is performed over a 3x3 window to reduce noise. Finally, the coordinates of each pixel are computed using triangulation.

Once a range map has been computed from the stereo imagery, we convert it into a voxel-based map representation. We first rotate the data such that it has the same relative orientation as the map we are comparing it to. Here we operate under the assumption that the orientation of the robot is known through sensors other than vision (for example, both Sojourner and Rocky 7 have a gyrocompass and accelerometers and Rocky i also uses a sun sensor for orientation determinat ion). However, these techniques can be generalized to determine the robot's orientation, as well.

The next step is to bin the range points in a three-dimensional occupancy map of the surroundings at some specified scale. We eliminate tile need to search over the possible translations of the robot in the z-direction by subtracting a local average of the terrain height froth each cell (i.e. a high-pass filter). This step is not strictly necessary, and it reduces our ability to determine height changes in the posit ion of the robot, but it also reduces the computation time that is required to perform localization. A subsequent step can be performed to determine the rover height, if desired. Each cell in the occupancy map that contains a range pixel is said to be *occupied*, and the others are said to be *unoccupied*.

Figure 2 gives an example of a terrain map that was generated using imagery from the Mars Pathfinder mission.

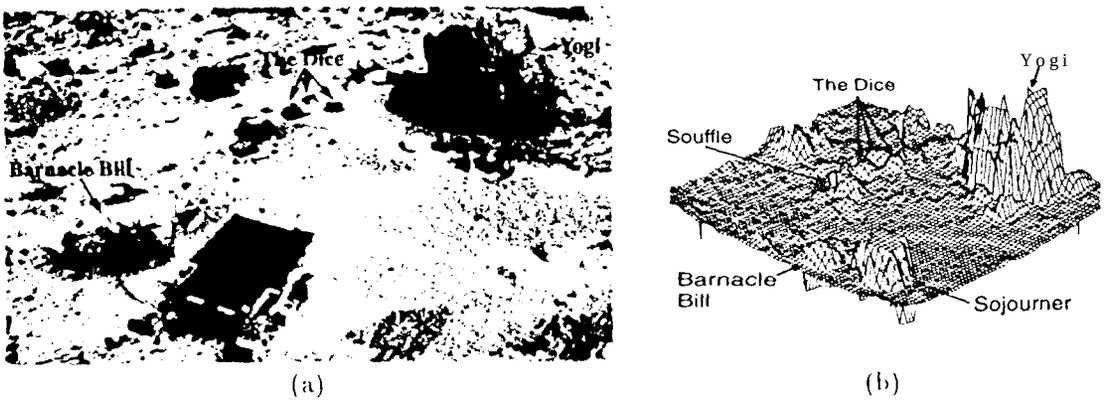


Figure 2: Terrain map generated from Pathfinder imagery. (a) Annotated image of Sojourner and rocks on Mars. (b) Terrain map generated from stereo imagery.

3 Map similarity measure

We have developed a similarity measure for comparing images and maps by reformulating a conventional image matching measure based on tile Hausdorff distance [5] in probabilistic terms using the principal of maximum likelihood est imation [9].

In order to formulate the matching problem in terms of maximum likelihood estimation, we must have some set of measurements that arc a function of the rover position. We use the distances from the occupied voxels in the local occupancy map to their closest voxels io the global map of the terrain with respect, to some relative position between the maps. Since we search for the best relative position between these maps, these distances are variables. Let us say that we have n occupied voxels io our local map. We denote the distances for these voxels at some position of the rover by $D_1 \ldots . D_n$. While these distances are not independent of each other, we model them as such. Recent work on determining the the probability of a false positive for matching with the Hausdorff distance [3, 10] provides support for treating these variables independently. We thus formulate the likelihood function for the rover position, X, as the product of the probability distributions of these distances:

$$L(X) = \prod_{i=1}^{n} p(D_i; X), \tag{1}$$

For convenience, we work in the $\ln L(X)$ domain:

$$\ln L(X) = \sum_{i=1}^{n} \ln p(D_i; X) \tag{2'}$$

The position yielding the maximum likelihood is taken to be the position of the rover. The probability distribution function (PDF) that is used for each voxel, $p(D_i; X)$, determines the matching measure that. is used between the occupancy maps. A simple two-valued PDF yields a measure equivalent to the Hausdorff fraction (which is a commonly used measure for image matching [9]):

$$\ln p(D_i; t) = \begin{cases} k_1 + k_2 & \text{if } D_i \leq \delta \\ k_1 & \text{otherwise} \end{cases} \tag{3}$$

The relative values of k_1 and k_2 are irrelevant (as long as $k_2 > 0$). In practice, we use $k_1 = 0$ and $k_2 = 1$.

Superior results can be achieved by modifying this probability distribution function [9]. Uncertainties inherent in the occupancy maps can be incorporated and we need not use this simple two-valued PDF. For example, we have found that a normal distribution with a constant additive term works well:

$$p(D_i; t) = \max_{g \in G} k_1 + k_2 e^{-\|t(l_i) - g\|^2 / k_3}, \tag{4}$$

where G is the set of occupied pixels in the global map, and $t(l_i)$ is the location of the ith occupied pixel in the local map according to some relative position, t, between the maps. This distribution models the case where the error in feature localization in the occupancy map has a normal distribution. The added constant allows for cases where features are not found at all (e.g. due to range shadows). Note that k_1, k_2, and k_3 are constants that should be set based on the error distribution in the location of the pixels and the probability of missing a feature,

4 Finding the most likely position

We locate the most likely rover position by adapting a multi-resolution search strategy that has been applied to conventional Hausdorff matching applications [10, 1 1].

As noted previously, we assume that the orientation of the rover is known from other sensors. Furthermore, the application of a high-pass filter to the occupancy maps removes the need to search in the z-direction to determine the position of the robot. We thus search only in the x- and y-directions. This two-dimensional pose space is discretized at the same resolution as the occupancy maps so that neighboring positions in the pose space move the relative positions of the maps by one map cell.

We first test the nominal position of the rover given by the dead-reckoning so that we have an initial position and likelihood to compare against. Next, the pose space is divided into rectilinear cells. Each cell is tested using, conservative testing techniques to determine whether it could contain a position that is better than the best found so far. Cells that cannot be pruned are divided into smaller cells, which are examined recursively. When a cell containing a single pose in the discretized pose space is reached, this pose is tested explicitly.

The key to this method is a quick method to test the cells. A cell is allowed to pass the test if it dots not contain a good pose, but it should never prune a cell that contains a pose that should pass the test, since this could result in the best position being missed.

To determine whether a cell C could contain a pose that is superior to the best one found so far according to the similarity measure described above, we examine the discrete pose c closest to the center of the cell. In order to place a bound on the Lest position within the cell, we compute the maximum distance between the location to which a cell in the local occupancy map is transformed into the global occupancy map by c and by any other pose in the cell. Denote this distance AC. If we treat poses as functions that transform positions in the local map into positions in the global map then A_c can be written:

$$\Delta_C = \max_{p \in C} \max_{l \in L} ||p(l) - c(l)||, \qquad (5)$$

where L is the set of occupied pixels in tile local map.

Let D_i^C denote the maximum likelihood that the ith occupied cell in the local map can achieve according to (4) with respect to any pow in cell C:

$$D_i^C \leq \max_{||x - c(l_i)|| < \Delta_C} \ln p(D_G(x); c), \qquad (6)$$

where l_i is the ith occupied cell in the local map and $D_G(x)$ is the distance transform of the global map. The distance transform yields the distance from any cell in the map to the closest occupied pixel in the map. This operation determines, for each occupied cell in the local map, the maximum likelihood that can be achieved over a radius of A_c from the relative position in the global map that the local cell is mapped by c. These values can often be calculated efficiently over the entire global map, depending on the L_p norm used [9].

A bound on the overall likelihood yielded by the best position in the cell is given by:

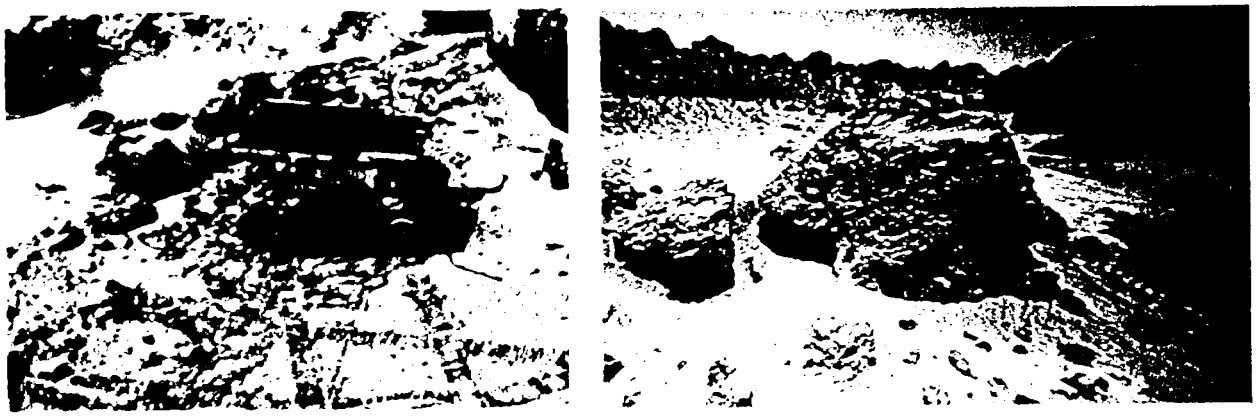
$$\max_{X \in C} L(X) \leq \sum_{i=1}^{n} D_i^C \qquad (7)$$

If this likelihood does not surpass the best that we have found so far, we can prune the entire cell from the search. Otherwise the cell is divided into two cells of the same size by slicing it along the longest axis and the process is repeated recursively until cells at the lowest level arc reached.

In order to implement this procedure efficiently, a breadttl-first search of the cell hierarchy is used. We maintain the invariant that each of the cells at the same level of the search tree nave the same dimensions and the breadth-first search examines all of the cells at each level before proceeding to the next level. Note that Δ_C is a function of only the dimensions of the cell, and not the position of the cell, when the pose space consists of translations. Furthermore, each D_i^C is a function of only Δ_C and the position that the center of the cell maps the local cell into the global map, so we can compute all of these values efficiently for each level of the search tree using a dynamic algorithm [9]. This allows each cell to be processed very quickly. For each occupied cell in the local map, we must only determine the position to which it is transformed into the global map, look up the appropriate D^C at this position and add it to the running total.

5 Results

We have tested these techniques using data from the Mars Pathfinder mission. A map of the terrain surrounding the Pathfinder lander was first generated using stereo imagery. For each position of Sojourner



(a) (b)

Figure 3: Sojourner on sol 21(near "Souffle"). (a) Composite image from the lander. (b) Image from Sojourner.

at which we tested the localization techniques, we generated au occupancy map of the terrain using range data from Sojourner's stereo cameras. This local map was then compared to the global map from the lander.

Unfortunately, this has only been possible at a few locations due to a limited amount of data returned to Earth, a lack of interesting terrain in some of the imagery we do have, aud a lack of ground truth for any positions except where Sojourner was imaged by tile lander cam eras. In practice, these techniques could he exercised much more frequently since they would uot require downlinking image data to Earth and the ground truth is only necessary for testing. We envision a scenario where the data from the rover's navigation cameras, which would be operating frequently iu order to perform obstacle detection, would be used to perform localization whenever sufficient terrain was evident in the imagery. In addition, the imagery from mast cameras could be used for localization] when the positional uncertainty grows beyond the desired level and the imagery from the navigation cameras is unsuitable.

As au example of the data, Figure 3 shows the position of Sojourner as seen from the lander and the view from Sojourner at the eud of sol 21[1] of the Mars Pathfinder mission. In addition, Figure 4 shows a subset of the digital terrain map of the site computed using lander imagery, which we used as the global map, and a digital terrain map computed from Sojourner imagery, which we used as tile local map for sol 21. Note that the stereo data obtained from Sojourner is uot as good as we hope to achieve in future missions. Accurate stereo data is achieved only for the central portion of the Sojourner imagery due to inaccurate calibration of the fish-eye lenses. The field-of-view that we have to work with is thus relatively small. However, we have achieved good localization results with this data.

'1'able 1 shows the results of localization using the techniques described in this paper versus the localization that was obtained by human operator through overlaying a rover model on the stereo data obtained from imaging the rover from the lauder. For sol 42, we have two localization results, oue prior to aud one after a turn by the rover. The operator localization was performed after tile turn.

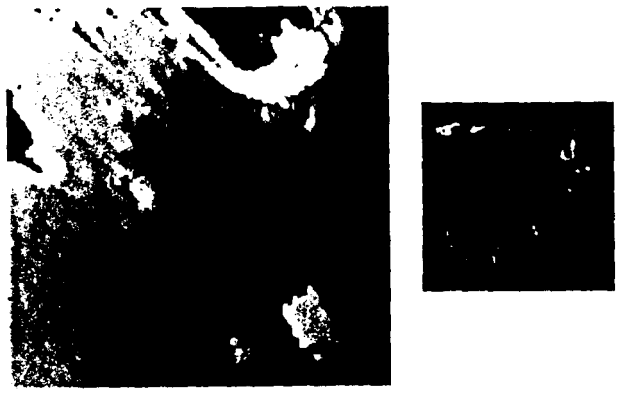
The results show' close agreement between our techniques and the oj)t'rater localization for four of the SOIS. For sois 4, 27, aud 72, there is some disagreement. Possible sources of error include inaccurate calibration of either the rover or lander cameras and operator error in performing localization. Manual examination of the maps indicates that the localization techniques determine the qualitatively correct position io these cases. We conclude that these techniques cau perform localization nearly as well as a human operator.

Note that these techniques require only a few seconds to perform localizatio1, both for these tests, which have been performed ou a work-station, aud in our implementation on-board Rocky 7.

6 Summary

We have described a method for performing rover self-localization by per forming maximum likelihood matching of terrain maps. We first generate a lo-

[1] A *sol* is a Martian day



(a) (b)

Figure 4: Digital terrain maps that were matched to localize Sojourner on sol 21. Black areas indicate no data. (a) Map of terrain near "Souffle" created using lander imagery (taken on sol 2). (b) Map of terrain in front of Sojourner from rover cameras (taken on sol 21).

cal map of the terrain using, stereo vision. This map is compared to a global map encompassing the same terrain to determine the optimal relative position between the maps using a maximum likelihood formulation of image matching techniques. This technique is guaranteed to find the best position in some discretization of the pose space and does not require an initial estimate of tile rover position.

The goal of this method is to provide greater autonomy for Mars rovers and we have applied these techniques to data from the Mars Pathfinder mission. While tbe data that we have is limited, and the quality is not as high as we expect in future missions, rover

	Operator		Localization	
sol	x	y	x	y
4	3.28	-2.69	3.01	-2.64
10	4.34	-3.24	4.24	-3.27
21	3.32	-2.60	3.37	-2.65
27	-5.42	2.85	-4.98	2.75
42a	-3.00	-1.86	-3.02	-1.87
42b	-3.00	-1.86	-3.00	-1.87
72	-8.93	-1.57	-8.99	1.35

Table 1: Comparison of rover positions determined by a **human** operator overlaying a rover model on stereo data of the rover and by the localization techniques described in this paper.

localization with accuracy of nearly the same quality as that obtained from a human operator has been demonstrated.

Areas that bear further study are tile development of a localizability measure for terrain maps in order to plan effective localization steps, and the development of a probabilistic uncertainty measure so that these techniques can be combined with other methods for performing localization. In the future, we plan to integrate these techniques into an organized navigation methodology, in which a Kalman filter is used to synthesize a rover position estimate from a variety of sensors and the rover's path planner interacts with the Kalman filter and localization techniques to plan when and where localization should be performed.

Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

This work is an element of the Long Range Science Rover project, which is developing technology for future Mars missions. This project is funded as part of the NASA Space Telerobotics Program, through the JPL Robotics and Mars Exploration 'J'ethnology Office.

References

[1] F. Cozman and E. Krotkov. Automatic mountain detection and pose estimation for teleoperation of *lunar rovers*. In *Proceedings of the IEEE Conference on Robotics and Automation*, volume 3, pages 2452-2457, 1997.

[2] 1). B. Gennery. [east-squares camera calibration including lens distortion and automatic editing of calibration points. In A. Grün and T. S. Huang, editors, *Calibration and Orientation of Cameras in Computer Vision*. Springer-Verlag, in press.

[3] W. E. 1,. Grimson and D. P. Huttenlocher. Analyzing the probability of a false alarm for the Hausdorff **distance** under translation. In *Proceedings of the Workshop on Performance versus Methodology in Computer Vision*, Pages 199 205, 1994.

[4] S. Hayati et al. The Rocky 7 rover: A Mars sciencecraft **prototype**, In *Proceedings of the IEEE Conference on Robotics and Automation*, volume 3, pages 2458 2464, 1997.

[5] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff **distance**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850-863, September 1993.

[6] L. Matthies. **Stereo** vision for planetary rovers: Stochastic modeling to **near** real-time implementation. *International Journal of Computer Vision*, 8(1):71-91, July 1992.

[7] L. Matthies, A. Kelly, 1. Litwin, and G. Tharp. Obstacle detection for unmanned ground vehicles: A progress report. In *Proceedings of the International Symposium on Robotics Research*, pages 475-486, 1996.

[s] L. H.Matthies, C. F. Olson, G. Tharp, and S. Laubach. Visual localization methods for Mars rovers using lander, rover, and descent imagery. In *Proceedings of the 4th International* Symposium on Artificial Intelligence, *Robotics and* Automation irr *Space*, pages 413–418, 1997.

[9] C. F. Olson. A probabilistic formulation for Hausdorff matching. Technical report, Jet Propulsion Laboratory, 1997.

[10] C. F. Olson and D. P. Huttenlocher. Automatic target recognition by matching oriented edge pixels. *IEEE Transactions* on image Processing, 6(1):103- 113, January 1997.

[11] W. J. Rucklidge. *Efficient Computation of the Minimum Hausdorff* Distance *for Visual Recognition.* PhD thesis, Cornell University, 1995.

[12] D. Shirley and J. Matijevic. Mars Pathfinder microrover. *Autonomous Robots*, 2:283–289, *1995.*

[13] R. Talluri and J. K. Aggarwal. Position estimation techniques for an autonomous mobile robot - A review. In C. H. Chen, L. F. Pau, and P. S. P. Wang, editors, *Handbook of Pattern* Recognition *and Computer Vision*, *chapter 4.4*, pages 769-8o1. World Scientific, 1'393.

[14] Y. Yakimovsky and R. Cunningham. A system for extracting three-dirneusional measurements from a *stereo* pair of TV cameras. *Computer* Vision, *Graphics, and Image Processing*, 7:195–210, 1978.